



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2014

Semantic role clustering: an empirical assessments of semantic role types in non-default case assignment

Bickel, Balthasar ; Zakharko, Taras ; Bierkandt, Lennart ; Witzlack-Makarevich, Alena

Abstract: This paper seeks to determine to what extent there is cross-linguistic evidence for postulating clusters of predicate-specific semantic roles such as experiencer, cognizer, possessor, etc. For this, we survey non-default case assignments in a sample of 141 languages and annotate the associated predicates for cross-linguistically recurrent semantic roles, such as ‘the one who feels cold’, ‘the one who eats sth.’, ‘the thing that is being eaten’. We then determine to what extent these roles are treated alike across languages, i.e. repeatedly grouped together under the same non-default case marker or under the same specific alternation with a non-default marker. Applying fuzzy cluster and NeighborNet algorithms to these data reveals cross-linguistic evidence for role clusters around experiencers, undergoers of body processes, and cognizers/perceivers in one- and two-place predicates; and around sources and transmitted speech in three-place predicates. No support emerges from non-default case assignment for any other role clusters that are traditionally assumed (e.g. for any distinctions among objects of two-argument predicates, or for distinctions between themes and instruments).

DOI: <https://doi.org/10.1075/sl.38.3.03bic>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-98913>

Journal Article

Originally published at:

Bickel, Balthasar; Zakharko, Taras; Bierkandt, Lennart; Witzlack-Makarevich, Alena (2014). Semantic role clustering: an empirical assessments of semantic role types in non-default case assignment. *Studies in Language*, 38(3):485 - 511.

DOI: <https://doi.org/10.1075/sl.38.3.03bic>

Semantic role clustering

An empirical assessment of semantic role types in non-default case assignment*

Balthasar Bickel*, Taras Zakharko*, Lennart Bierkandt† and
Alena Witzlack-Makarevich‡

*University of Zurich / †University of Jena / ‡University of Kiel

This paper seeks to determine to what extent there is cross-linguistic evidence for postulating clusters of predicate-specific semantic roles such as experiencer, cognizer, possessor, etc. For this, we survey non-default case assignments in a sample of 141 languages and annotate the associated predicates for cross-linguistically recurrent semantic roles, such as ‘the one who feels cold’, ‘the one who eats sth.’, ‘the thing that is being eaten’. We then determine to what extent these roles are treated alike across languages, i.e. repeatedly grouped together under the same non-default case marker or under the same specific alternation with a non-default marker. Applying fuzzy cluster and NeighborNet algorithms to these data reveals cross-linguistic evidence for role clusters around experiencers, undergoers of body processes, and cognizers/perceivers in one- and two-place predicates; and around sources and transmitted speech in three-place predicates. No support emerges from non-default case assignment for any other role clusters that are traditionally assumed (e.g. for any distinctions among objects of two-argument predicates, or for distinctions between themes and instruments).

1. Introduction

Apart from default or canonical case assignments, such as the assignment of accusative case to the most patient-like argument of transitives, many and perhaps most languages show alternatives in the form of non-default or non-canonical assignments for specific sets of predicates, e.g. the accusative on arguments of experience-denoting intransitives (as in German *mich friert* ‘I am cold’). It is often assumed that non-default marking of this kind does not occur at random. Indeed, several hypotheses and theories have been put forward that seek to predict the way in which semantic types of predicates associate with non-default case

assignments across languages (e.g. Tsunoda 1985, 2004, Onishi 2001, Haspelmath 2001, Malchukov 2005). However, all these hypotheses assume that predicates with non-default case assignments fall into natural semantic types such as EXPERIENCE, MOTION, UNCONTROLLED EVENT, etc. In other words, it is assumed that all lexical tokens of predicates, i.e. items like the German verb *frieren* or the English predicate expression *be cold*, can be successfully mapped into more general and more abstract classes like EXPERIENCE. A prominent correlate of this assumption is that predicate-specific argument roles cluster into more general and more abstract argument types — “role complexes”, as the editors of this special issue call them — such as EXPERIENCER, THEME, or INSTRUMENT etc., and that these clusters are significantly similar across languages.

This assumption is controversial. It is usually debated in terms of a choice between theoretical frameworks, e.g. by appeal to the efficiency and elegance in describing general patterns of case assignment (e.g. capturing which intransitive verbs assign accusative rather than nominative case in German) or constructional constraints beyond case (e.g. in auxiliary choice or participle formation). Some theories assume role types (e.g. Lexical-Functional Grammar; Bresnan & Kanerva 1989, Butt 2008, Dalrymple & Nikolaeva 2011), others reject them (e.g. Role and Reference Grammar; Van Valin & Wilkins 1996, Van Valin 2005). In this paper we want to turn the debate into an empirical one. Based on a typological database, we assess the empirical evidence for role clusters, asking to what extent non-default case assignment suggests natural and cross-linguistically relevant clusters: is there cross-linguistic evidence that non-default case assignment indeed systematically carves out, say, experiencers and themes as general types among the sole argument of one-place predicates? Is there evidence for carving out, say, perceiver and agent types among the most agent-like argument of two-place predicates (such as *see* vs. *hit*, etc.)? Is there any evidence for such type distinctions as recipients vs. spatial goals among the non-moving argument of three-place predicates (e.g. *give* vs. *put*)?

We start by annotating case frames for predicate-specific roles (e.g. ‘the one who feels cold’, ‘the one who sees sth.’, ‘the one that gets hit by so’, etc.) that recur across languages and that can be reasonably identified by translational approximation. We limit our attention to non-default case assignment (and non-default case alternations), assuming that defaults have no semantic specification of their own and cover everything that is not covered by non-default cases (or alternations). We then examine on a typological database to what extent predicate-specific semantic roles are grouped together by the same non-default cases (or alternations) in each language and derive from this a measure of dissimilarity of the roles across languages. The resulting dissimilarity matrix is then mined for statistical clusters, applying algorithms for fuzzy clustering (Kaufman & Rousseeuw 1990) and NeighborNet (Bryant & Moulton 2004, Huson & Bryant 2006) analysis. Any

resulting cluster of predicate-specific roles is potentially indicative of cross-linguistically relevant role complexes.

In the following, we first explain our notions of non-default case assignment and generalized argument classes (Section 2). Section 3 explains our database and the way we developed the cross-linguistic annotations of predicate-specific semantic roles. Section 4 describes the data-mining algorithms we used. Results of these are then presented in Section 5 and discussed in Section 6 in the light of expectations from the literature. The final section summarizes our findings.

2. Non-default case assignment and generalized argument classes

Many languages exhibit diverse possibilities of case assignment.¹ This can be illustrated with the following examples from Chechen (ISO639.3:che; Nakh-Daghestanian; Zarina Molochieva, p.c.). The clauses in (1) show that the sole argument of a one-place predicate can be in the absolutive, the dative, the ergative, and the allative case. (2) shows a selection of possibilities available for the arguments of two-argument predicates:²

- (1) a. *so ohw-v-uuzhu-u.*
1S.ABS down-V-fall-PRS
'I fall down.'
- b. *suuna j-ouxa j-u.*
1S.DAT J-hot J-be.PRS
'I am hot.'
- c. *as jouxarsh tyxyi-ra.*
1S.ERG cough hit-WITNESSED.PST
'I was coughing.'
- d. *soega nir qiet-a.*
1S-ALL diarrhea strike-PRF
'I've got diarrhea.'
- (2) a. *as wazh b-u'-u.*
1S.ERG apple(B).ABS B-eat-PRS
'I eat apples.'
- b. *so hwo-x taxana qiet-a.*
1S.ABS 2S-LAT today meet-PRS
'I meet you today.'
- c. *suuna Zaara j-iez-a.*
1S.DAT Zara(J).ABS J-love-PRS
'I love Zara.'

Clearly, case assignment is sometimes not an isolated phenomenon but is part of a larger constructional choice. In (1d), for example, the allative is conditioned by the fact that the predicate is not expressed by a simple stem but instead by a complex lexicalized expression that involves the allative-assigning verb stem *qiett*- ‘strike’. Strictly speaking, then, the lexical entry ‘have diarrhea’ associates with the entire complex construction ‘allative+*qiett*-’ and not just with the allative declension form. In this paper, we gloss over this complication for the following reason: Our interest lies in whether or not the roles that are licensed by various cross-linguistically identifiable lexical meanings (such as the role of the single argument of ‘have diarrhea’) are treated alike or not in a language, and whether there are systematic patterns behind this treatment across languages. For this question, it does not matter if a specific predicate meaning associates with a simple case choice or with a complex constructional choice of case and complex predicate structure at the same time. This difference is as irrelevant to our question as the difference between a case choice that affects only a simple suffix and one that involves some complex expression consisting of, say, a declension form and an adposition.³ As we will explain further below, we base our analysis of roles exclusively on the semantics of lexical entries (where Chechen *nir qiett*- licenses a single argument S just like English ‘have diarrhea’) and not on the formal shape of these entries (where one can argue about the transitivity of the expressions).

It is commonly assumed that some types of case assignment represent the basic or canonical choice and others a non-basic, non-canonical choice. In Chechen, for example, one would consider the absolutive in (1a) and the ergative-absolutive frame in (2a) to be the basic choices. The range of individual predicates in basic case frames is typically open-ended, with no specified semantic limits. Open-ended classes of this kind are difficult to survey across languages because sufficiently rich dictionaries are scarce.

One way out of this problem is to proceed with an *a priori* list of universal predicate meanings (like ‘eat’, ‘have diarrhea’, etc.) whose case assignments can then be catalogued for every language.⁴ Like all onomasiological (denotation-based, stimuli-based) approaches, this procedure allows easy comparison, but the pre-selection of predicate meanings brings with it the risk that the results are in part pre-determined. For example, it makes a difference for role clusters among intransitives how many different verbs of body functions (e.g. ‘belch’) there are, how many experience-related verbs (e.g. ‘be cold’) there are, etc.: if there are more experience-related meanings than body-function meanings in a list, evidence from case assignment patterns related to experiences weighs more than evidence related to body functions in cluster analyses, and this artificially favors the detection of experiencer clusters over role clusters related to body functions.

While these problems can be kept somewhat at bay by enlarging lists of surveyed meanings and by trying to avoid Eurocentrism when compiling them, we explore an alternative approach here, concentrating exclusively on non-basic case assignment. The predicates associated with non-basic case assignment have the advantage that they are positively characterized by lists of verbs (which are also typically retrievable in grammars because one needs to say when the relevant non-basic cases show up). Basic case assignment patterns, by contrast, can be expected to apply to open-ended lists of verbs, with an equally open diversity of meanings. Lists of verbs assigning non-basic cases can be readily surveyed and compared without any *a priori* assumptions about what to expect. But this approach is not without problems either. The most pressing one is how one can in fact distinguish basic from non-basic case frames.

There are basically two classes of approaches to this, each of them replacing the intuitive notion of a basic choice in case assignment by more concrete concepts that can be better operationalized. In one approach, the notion of a basic choice is replaced by that of canonical arguments, or, more precisely, notions of canonically intransitive, canonically transitive, and canonically ditransitive argument frames. Canonicity is in turn established on the basis of a range of morphosyntactic, or semantically-grounded morphosyntactic, criteria so that, for instance, only the frames with accusatively-marked objects, or only with accusatively-marked and affected patient objects, or only with objects which can be promoted to subjects through passivization and which denote affected patients, are considered canonically transitive (cf. e.g. Onishi 2001, relying on Dixon 1994, but also more generally any research relying on notions like “quirky subjects”, “oblique objects”, etc.). In another approach, the notion of basic choice in case assignment patterns is grounded in prototypical predicate meanings. The choice is then based on pre-established notions of what would be prototypical representatives of one-, two-, and three-argument predicates, e.g. one would define predicates meaning something like ‘kill’ or ‘break’ as the prototypical representatives of two-place predicates (in the spirit of Comrie 1981) and take the case frame of these predicates to be basic.

The first approach has been criticized for mixing semantic and syntactic criteria that are not strictly comparable across languages (Haspelmath 2011), and we do not adopt this approach here for this reason. The second approach does not suit the purpose of our investigation because it builds into a theoretical assumption what we want to explore empirically: the approach assumes *a priori* that across languages, predicates fall into at least two basic semantic types or clusters, a prototypical one including (in the case of two-argument predicates) meanings like ‘kill’ or ‘break’, and a non-prototypical one including meanings like ‘love’ or ‘see’. This may well be the case, but if so, we expect it to emerge empirically from a cluster analysis.

Therefore, we need an alternative approach: we approach the notion of basic case frames with the idea of DEFAULT case frames: case frames that are assigned when there is no other case specification in the lexical entry of a predicate. Default case frames in this sense are expected to be licensed by predicates that form an open class, and thereby by whatever class has the largest number of members in the lexicon and is most productive.⁵ For Chechen, this criterion establishes the predicate classes in (1a) and (2a) as the largest and most productive ones, and therefore their case frames, i.e. $\langle \text{ABS} \rangle$ and $\langle \text{ERG}, \text{ABS} \rangle$, as the default. Any case assignment that deviates from these patterns is considered as the non-default frame of one-argument and two-argument predicates, respectively. For non-default frames, one needs to specify case assignments in the lexical entry of the predicate, e.g. one needs to stipulate that Chechen *qiet-* ‘meet’ assigns an $\langle \text{ABS}, \text{LAT} \rangle$ frame (as shown by (2b) above).

This approach allows uncontroversial identification of default and non-default case frames among one-argument and two-argument predicates. The situation might at first sight seem less straightforward for split-S or “active” languages, which according to some analyses possess two equally substantial and equally salient sets of one-argument predicates (e.g. Comrie 2005). However, in each language of our sample, one of the one-argument classes clearly dominates the other in terms of the number of predicates.⁶ Problems arise, however, with three-argument predicates. As Malchukov et al. (2010) observe, three-argument predicates tend to be substantially less frequent in the lexicon than other predicates, and the small sets one finds are often very heterogeneous. In line with this observation, many languages in our sample have no clear default class of three-argument predicates. This is so, for example, in Tsamai (ISO639.3:tsb; Cushitic; Sava 2005). There are two case frames for three-argument predicates — $\langle \text{NOM}, \text{ACC}, \text{LOC} \rangle$ and $\langle \text{NOM}, \text{ACC}, \text{ABL} \rangle$ — and none of them outranks the other to such an extent that it could be taken to be the default choice. In such cases, we treated all classes as if they were non-defaults.

Once non-default case assignment is identified in each language, the question arises of how one can compare the way in which non-default case assignment is associated with predicate-specific semantic roles. It would not be helpful to compare cases across just any kind of predicate-specific semantic roles because there is a fundamental difference between, say, the two roles licensed by two-place predicates and the three roles licensed by three-place predicates. For example, one would expect more predicate-specific semantic roles related to spatial transfer in three-place predicates than in two-place predicates, and this already imposes a limit on the kinds of clusters that can be found. Also, random comparisons of roles across predicates can create patterns that yield no real insight: if, say, the ‘what is eaten’ argument of ‘eat’ and the ‘one who meets’ argument of ‘meet’ happen to be

assigned the same case in a language (e.g. absolutive, as in Chechen in (2)), one would not infer that these roles have something in common in their semantics. Therefore, comparisons of predicate-specific semantic roles are best based on a *tertium comparationis* that keeps general classes of arguments, such as the two arguments of ‘eat’, distinct.

Traditionally, such classes are kept distinct on the basis of grammatical relations, and role clusters are then sought separately within ‘subjects’ and within ‘objects’. However, since grammatical relations incur well-known problems of cross-linguistic comparability due to their language-specific and construction-specific nature (cf. Dryer 1997, Croft 2001, Bickel 2011, among many others), we proceed in a different way. We distinguish between argument classes first by numerical valence: the sole argument of one-argument predicates, the two arguments of two-argument predicates, and the three arguments of three-argument predicates. The sole argument of one-argument predicates is symbolized as S. The arguments of two- and three-argument predicates are then distinguished on the basis of cross-linguistically viable lexical entailment properties (following a line of research going back to Dowty 1991):⁷

- (3) Lexical entailments defining generalized argument classes
 - a. A vs. P: A accumulates more lexical entailments than P on the following properties:
 - causing an event (e.g. *A hits P*, *A kisses P*, *A goes to P*, *A meets P*)
 - volitional (e.g. *A hits P*, *A kisses P*)
 - sentient (e.g. *A sees P*, *A looks at P*, *A loves P*, *P pleases A*)
 - independently existing (e.g. *A bakes P*, *A makes P*)
 - having possession over another participant (e.g. *A has P*, *P belongs to A*)
 - b. G vs. T: G accumulates more lexical entailments than T on the following properties:
 - stationary relative to movement of another participant (e.g. *A gives T to G*, *A loads T onto G*, *A covers G with T*, *A cuts G with T*)
 - receiving or being exposed to an experience (e.g. *A shows T to G*, *A tells T to G*)

For instance, *kiss* — as in *Lisa kissed Tom* — entails that *Lisa* is causing the event of kissing, behaves volitionally, is sentient of this event, and exists independently of kissing. *Tom* exists independently of kissing, but he is neither causing this event, nor is he behaving volitionally, nor is he necessarily sentient of it (e.g. if he is in coma or asleep). Thus, the kisser accumulates more of the lexical entailments defining the A argument than the kissee and qualifies as the A argument, while the kissee is then necessarily the P argument.

Note that 'A' in (3) stands for the A argument class of two-argument predicates only. Three-argument predicates have an 'A₂' argument (Bickel & Nichols 2009, Bickel 2011), and this is distinguished from T and G in the same way as A is distinguished from P arguments. The main reason for distinguishing between A and A₂ is that the kinds of meanings covered by the A₂ role of three-argument predicates are different from those of two-argument predicates (compare e.g. 'the one who sends', 'the one who gives' among three-argument predicates with 'the one who perceives', 'the one who hits' etc. among two-argument predicates). This difference often leaves no reflex in case marking and most languages treat A₂ in exactly the same way as the A argument of two-argument predicates (but see Bickel & Nichols 2009). However, with non-default cases, the difference between A and A₂ matters more commonly. For example, emotion-related dative experiencers ('be afraid of something or someone') are typically not found among three-argument predicates but only among two-argument predicates.

Note that the generalized argument classes, as we define them here, are strictly independent of their morphosyntactic realization and thereby of any association with a specific case frame (or indeed, of any association with a case frame plus some formal sub-structure in the predicate, as discussed in Section 2). This makes it possible for these classes to serve as a *tertium comparationis* when comparing the cases that specific predicates associate with. Applying our definitions of generalized classes to the Chechen examples above, we obtain three non-default case frames for the data in (1), viz. ⟨S-DAT⟩, ⟨S-ERG⟩, and ⟨S-ALL⟩, and two non-default case frames for the data in (2), viz. ⟨A-ABS, P-LAT⟩, and ⟨A-DAT, P-ABS⟩.

While they are fully independent of morphosyntax, the argument classes in (3) are not fully independent of semantic roles and role clusters: they presuppose that agents and patients, or goals and themes are different from each other, respectively, and could never end up in the same cluster. This is a limitation to our approach. However, the damage is minimal because, as noted above, clusters combining agents and patients, or goals and themes, would be difficult, perhaps even impossible, to interpret semantically, i.e. they would not yield a consistent role type. Also, the way we set up the classes still leaves much room for role clusters within each class, all of them with a realistic chance for the kind of semantic homogeneity that theories of semantic roles predict, e.g. an experiencer vs. a theme role among S and A arguments, or a recipient vs. spatial goal role among G arguments.

Equipped with argument classes and a definition of non-default case assignment, we can now compare case assignment patterns for each generalized class: are there any cross-linguistic trends in how non-default case assignments cluster predicate-specific roles in the class of S arguments, in the class of A arguments, T arguments, etc.?

3. Data

We surveyed a sample of 141 languages and coded their case systems. The sample is geographically widespread, covering the whole world (see the map in Figure 1).⁸ The sample is also genealogically balanced along the lines proposed by Dryer (1989): we excluded non-default case assignment rules that exactly replicate rules in other languages within the same major branch of a family (in the sense of Nichols et al. 2013). This rules out signals being occasionally inflated by larger families, but it also means that our study is only exploratory at this point and cannot assess the diachronic dynamics that lead to the patterns we find.⁹

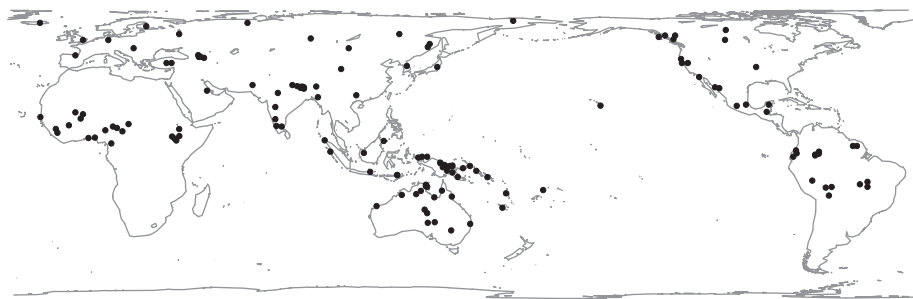


Figure 1. Languages in the sample ($N = 141$)

In each language of the sample, we identified each set of predicates that is associated with a specific non-default case frame, or a specific case frame alternation involving a non-default marker (such as alternations conditioned by tense or aspect or by referential properties in the form of differential object marking, or by perspectivization choices as in ‘spray/load’-type alternations etc.). We annotated each resulting set for the range of meanings in its members, approximated by English translation equivalents as found in or inferred from available lexical resources in dictionaries and grammars. We only included meanings that are specific to a predicate: for example, when the description of a non-default case frame only vaguely refers to “verbs of transfer”, without further specifications (e.g. whether this includes verbs like ‘send’ or ‘tell’), we did not include this as the meaning ‘transfer’.

There is no doubt that our data collection and annotation misses relevant predicates in some languages, e.g. because the predicates were simply forgotten when grammar authors set up lists. Translations can include errors as well. Specifically, we, or our sources, may have occasionally missed semantic compositionality and inadvertently analyzed as a single complex predicate (e.g. ‘cough’) what really should be analyzed as a two-place predicate (‘throw out particles’, with particles as a referring expression).¹⁰ As a result, it is clear that our data is noisy to some extent. However, if semantic role clusters are cross-linguistically real and

Table 1. Chechen non-default case frames or case frame alternations and the meanings of the associated predicates. (The ERG/ABS-alternation is conditioned by aspect; see Molochieva 2010).

Case frame/alternation	Meanings
⟨S-DAT⟩	have fever, feel hot, feel good
⟨S-ERG/ABS⟩	blink, lose hope, be insulted, cough, sin, grow
⟨S-ALL⟩	start coughing, have diarrhea
⟨A-DAT, P-ABS⟩	see, hear, like, love, remember, lack
⟨A-ABS, P-ALL⟩	look at, wait for, scold
⟨A-ABS, P-LAT⟩	depend on be proud of, fear
⟨A-ERG/ABS, P-DAT⟩	look after, surround, lock, help
⟨A-ERG/ABS, P-LAT⟩	abuse, forgive
⟨A ₂ -ERG/ABS, T-ABS, G-ALL⟩	say, send, loan, ask, remind
⟨A ₂ -ERG/ABS, T-LAT, G-ABS⟩	load, support
⟨A ₂ -ERG/ABS, T-INS, G-ABS⟩	fill, smear, cover, congratulate

relevant, we expect them to leave statistical signals in our data despite all this noise — perhaps not very strong and clear, but at least detectable signals. At any rate, the raw data that we obtained in this way consists of tables as in Table 1 for Chechen.

Based on tables like these, we extracted the meanings associated with each CLASS-CASE COMBINATION, i.e. with each combination of an argument class (e.g. S) with a specific case or case alternation (e.g. DAT or ERG/ABS). For example, for the class of S arguments we would collect all predicate-specific roles covered by the same case or case alternation. Applied to Table 1, this results in one set including the roles of ‘the one who has a fever’, ‘the one who feels hot’, and ‘the one who feels good’ (covered by the class-case combination S-DAT), another set containing the roles of ‘the one who blinks’, ‘the one who loses hope’, ‘the one who feels insulted’, ‘the one who coughs’, ‘the one who sins’, and ‘the one who grows’ (all assigned ergative case in the continuous and absolutive in the non-continuous aspect, i.e. the class-case combination S-ERG/ABS), and finally, a set consisting of ‘the one who starts coughing’ and ‘the one who has diarrhea’ (the class-case combination S-ALL). To keep the description manageable we refer to the predicate-specific roles in the following simply by the argument class and the predicate meanings, e.g. ‘S-have_fever’ stands for an S-class semantic role ‘the one who has a fever’, etc.

The extraction of predicate-specific semantic roles with non-default case assignments is trivial for S-case combinations because each such combination is associated with a unique set of roles. For other arguments, the semantic roles need to be collected across case frames. For example, in Table 1, the absolutive-marked A occurs

in two different case frames, viz. $\langle A\text{-ABS}, P\text{-ALL} \rangle$ and $\langle A\text{-ABS}, P\text{-LAT} \rangle$. Therefore the class-case A-ABS is associated with the set union of the meanings of these two frames: {A-look_at, A-wait_for, A-scold, A-depend_on, A-be_proud_of, A-fear}.

When a class-case combination reflects the default, we do not analyze it further. For example, A-ERG/ABS in Chechen is the case alternation pattern that is associated with the default frame (cf. the data in (2a) above), and so we do not analyze the meanings covered by the class-case combination A-ERG/ABS. As a result of this, the meanings associated with the frames $\langle A\text{-ERG/ABS}, P\text{-DAT} \rangle$ and $\langle A\text{-ERG/ABS}, P\text{-LAT} \rangle$ in Table 1 enter the analysis only with regard to the P-DAT and P-LAT class-case combinations, where they do not correspond to a default. This results in the role sets {P-look_after, P-surround, P-lock, P-help} as covered by the dative case and {P-depend_on, P-be_proud_of, P-fear, P-abuse, P-forgive} as covered by the lative case.

A special situation arises when a language lacks a default case frame, which is (as noted earlier) fairly common among three-argument predicates. In these cases, we analyzed a class-case combination as a relevant (i.e. non-default) instance if (and only if) it does not extend across all case frames in three-argument predicates. We noted above (Section 2) the situation in Tsamai. In the two three-argument frames — $\langle A_2\text{-NOM}, T\text{-ACC}, G\text{-LOC} \rangle$ and $\langle A_2\text{-NOM}, T\text{-ACC}, G\text{-ABL} \rangle$ — the class-case combinations $A_2\text{-NOM}$ and $T\text{-ACC}$ occur in both case frames, and so we exclude these two, assuming that they cannot give evidence of any semantic specification. Thus, Tsamai three-argument predicates enter the analysis only in terms of the G class, with specifications for the class-case combinations G-LOC and G-ABL.

Once these analyses were performed, we searched through all sets of predicate-specific semantic roles that are defined by non-default case assignment and looked for recurrent items. We interpreted these items as proxies of cross-linguistically relevant roles (such as S-have_fever, A-love, P-fear, etc.). The question arises of what one would take as the minimum amount of recurrence for exploring cross-linguistic patterns in roles. Table 2 surveys how often roles recur across languages, i.e. how often they are mentioned in our database as being assigned a non-default case or undergo non-default case alternations.

Given the skewing towards low N in Table 2, any cut-off point will have to be relatively low unless one wants to throw out most of the data. Since any choice here would be arbitrary (say, requiring a role to be mentioned in at least 5 languages), we decided to take into consideration all roles that recur at least once (i.e. that are mentioned in at least two rules of non-default case assignment or non-default case alternation, $N \geq 2$). This means that even a single recurrence of a pattern (e.g. that two languages assign S-have_fever and S-be_cold the same non-default case) is taken as a relevant signal for role clustering. However, we will take the amount of recurrence into account when interpreting results.

Table 2. Number of predicate-specific roles that are assigned a non-default case or a non-default case alternation N times

	$N=1$	$N=2$	$N=[3,4]$	$N\geq 5$	Total
S	61	16	16	9	102
A	88	33	18	24	163
P	116	57	46	70	289
A_2	1	0	0	0	1
G	34	19	16	18	87
T	20	13	9	9	51

Note that for the A_2 class of arguments (in three-place predicates), only one language in our dataset shows a non-default class-case combination. This is Marathi (ISO639.3:mar; Indo-Aryan), where verbs of telling assign nominative instead of ergative in all tenses, whereas the default rule is to assign ergative in the past tense (Pandharipande 1997: 132). All other languages in our database assign A_2 the same case throughout, across all three-place predicates. We therefore exclude the class of A_2 arguments from further analyses.

Requiring roles to recur at least once (i.e. $N\geq 2$ in Table 2) leaves us with 41 predicate-specific roles in the class of S arguments, 75 in the class of A arguments (in two-place predicates), 173 in the class of P, 53 in the class of G, and 31 in the class of T arguments.

4. Methods

Our goal is to identify the extent to which non-default case assignment targets similar or dissimilar predicate-specific semantic roles. For example, if non-default case assignment to S recurrently treated roles related to experiences (S-feel_cold, S-feel_sick, etc.) alike, we would interpret this as evidence for a cluster of EXPERIENCER roles in the class of S arguments.

In order to derive a suitable measure for assessing whether semantic roles are treated similarly vs. differently, we first tabulate all roles and specify whether or not a given role is listed for each class-case combination in the database. Table 3 illustrates this for a selection of class-case combinations. For example, we observed above that the Chechen A-ABS combination associates with the set of predicate-specific semantic roles {A-look_at, A-wait_for, A-scold, A-depend_on, A-be_proud_of, A-fear}. In Table 3, this is represented as follows. Chechen A-ABS (in the first column) gets 0 (absence) for A-love and A-hate since these semantic roles are not in the set, but 1 (presence) for A-fear and A-wait_for because these

Table 3. Matrix of semantic roles

	Chechen	Djambarrpuyqu	Tsez	Nias	...
	A-ABS	A-NOM	A-DAT	A-ABS	...
A-love	0	1	1	1	...
A-hate	0	0	1	1	...
A-fear	1	1	0	1	...
A-wait_for	1	0	0	0	...
...

components are in the set. Other class-case combinations in Chechen and other languages show a different pattern.

We then compute the pair-wise mean dissimilarity across rows in the matrix, using the Jaccard distance. The Jaccard distance is defined as the proportion of differences (i.e. with ‘0’ in one row and ‘1’ in the other) among all columns that do not contain ‘0’ in both rows. Discarding pairs with both ‘0’ (or blanks) is important because the absence of some non-default case assignment only means that the roles are ignored by the relevant assignment rule, not that they are treated alike. Table 4 shows the dissimilarity matrix that is derived from Table 3 by computing the Jaccard distance.

Table 4. Dissimilarity matrix of semantic roles

	A-love	A-hate	A-fear
A-hate	1/3 = .33		
A-fear	2/4 = .5	3/4 = .75	
A-wait_for	4/4 = 1	3/3 = 1	2/3 = .66

For example, comparing the rows with A-love and those with A-hate in Table 3 shows three columns with pairs that are not both zeros, i.e. the Djambarrpuyqu, Tsez, and the Nias column. One of these shows presence (1) for A-love and absence (0) for A-hate; two show presence in both. This results in a Jaccard distance of 1/3. In the comparison of the rows with A-love and A-fear, four (i.e. here, all) columns have at least one ‘1’ in both rows. In two columns (Nias and Chechen) values disagree and so the distance is 2/4. We compute such dissimilarity matrices separately for each of the five argument classes retained for analysis: S, A, P, G, and T. The resulting matrices are then mined for clusters. There are many techniques available for this, each with its advantages and disadvantages (cf. Kaufman & Rousseeuw 1990, Rokach 2010). In order to increase the chances of

detecting patterns in the data we combined two methods: fuzzy cluster analysis and NeighborNets.

The basic idea of fuzzy cluster analysis is not to partition the data, here predicate-specific semantic roles, into a categorical set of clusters, but to partition the data into clusters with fuzzy boundaries and centered around a prototype. This fits with common assumptions that, like other semantic categories, semantic roles are structured around prototypes. Fuzzy cluster analysis assigns each member of a cluster a membership coefficient between 0 and 1, indicating the probability of being a member in this cluster (so that k represents equiprobable membership in k clusters). The challenge is to optimize these membership coefficients in such a way that the coefficients minimize within-cluster dissimilarity and maximize between-cluster dissimilarity, given a number of clusters. Kaufman & Rousseeuw (1990) introduced an iterative algorithm to solve this optimization problem (implemented in R (R Development Core Team 2013) and made available by Maechler et al. 2005).

As individual datasets allow for many ways of fuzzy clustering, one needs to determine the most adequate number of clusters (k). This is done by considering a reasonable range of clusters and selecting the one for further analysis that shows the clearest, “crispest” partition of items (here, semantic roles). The standard goodness-of-fit statistic for crispness of clusters is known as the Dunn statistic (or Dunn coefficient) which in its normalized form ranges from 1 (crisp, well-motivated clustering) to 0 (no evidence for clustering, i.e. membership is always equiprobable across all clusters). The best solution is then defined as the one that combines the lowest number of clusters (k) with the highest Dunn statistic. We search for this optimum ‘by hand’, i.e. by inspecting plots of Dunn statistics against numbers of clusters (printed below, in the Results section) and choosing the optimum as the point where the Dunn statistic reaches its first peak or where it begins to flatten out (i.e. where higher numbers of clusters no longer improve the statistic). Crisper clustering can generally be obtained by lowering what is called the membership exponent (ME), which serves to weigh the membership coefficients during the algorithm. While yielding crisper signals, low exponents can impede the convergence of the algorithm (Kaufman & Rousseeuw 1990). We therefore chose the minimal exponent that allowed convergence across datasets.

Fuzzy cluster analysis results in lists of sets and estimates of the degree to which each semantic role belongs to any one of these sets, with the prototype members having membership coefficients close to 1. What is missing from this method is a direct and easy-to-visualize estimate of the similarities and dissimilarities *between* clusters. These (dis-)similarities are of interest for our study because they could reveal higher-order relations between role clusters. In order to derive estimates of cross-cluster relations, we also performed splits-graph analyses of the data, using the NeighborNet method as implemented in the SplitsTree package (Bryant & Moulton 2004, Huson

& Bryant 2006). NeighborNet is an algorithm that solves a fundamental problem of visualizing dissimilarity (or distance) matrices like the one in Table 4: if we know the distances between 3 elements, we can draw the 3 locations on a plane without distorting the distances. If there are more elements and the distances vary freely, this requires more dimensions, defeating visualization on a plane. The NeighborNet algorithm basically solves the problem by splitting the paths between elements. We illustrate the application of this method on a small set of data in Table 4.

The distance matrix in Table 4 cannot be plotted on a plane without distorting distances: it is geometrically impossible to place A-wait_for in such a way that it has a distance of 1 to A-love and A-hate, and that at the same time the fourth item A-fear has a distance of one half to A-love and three quarters to A-hate. The NeighborNet in Figure 2 solves the problem by splitting up paths in such a way that they represent the distances faithfully. This can be verified by adding path lengths and comparing them to the original matrix, allowing for rounding effects. For instance, to find the distance between A-wait_for and A-fear, one calculates the sum of the lengths of the edges in the path between these two elements: $.457 + .125 + .084 = .666$. This sum corresponds to the Jaccard distance in Table 4.

We produced NeighborNet representations of the dissimilarity matrices of each argument class and then superimposed the results of the fuzzy cluster analysis to display where the prototypes are of each cluster and how far away individual components are from the prototype. For this we use different colors for each cluster and let the saturation of the color show the membership coefficient, starting with full saturation for a membership coefficient of 1 and then shading into less saturated colors and reaching white where membership coefficients equal equiprobable membership ($1/k$).

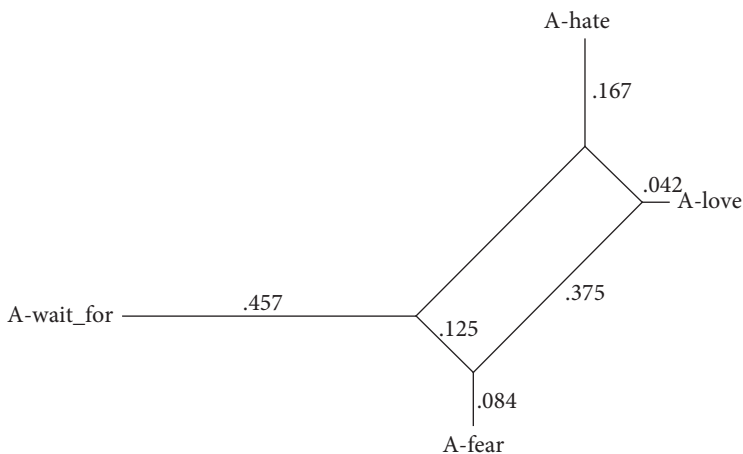


Figure 2. A NeighborNet visualization of the dissimilarity matrix in Table 4

Not all cluster signals have the same weight in interpretation: if two roles are frequently treated alike by non-default case assignment, this is stronger evidence for the cluster than if the (dis)similarity assessment is based on fewer non-default case assignment rules because the roles are marked by default cases in most languages. In order to keep this information visible when interpreting the analyses, we scaled the font sizes according to the frequency range of roles displayed above in Table 2: if roles are assigned non-default cases in only two languages, we choose the smallest font, if they are assigned non-default cases in between three and five languages, the role label is printed in the second-smallest size, and if they occur in between six and ten languages, the second-largest font issued. The largest font size is reserved for roles that are assigned non-default cases in at least 11 languages.¹¹

5. Results

With a membership exponent of $ME = 1.1$, it was possible to obtain fuzzy cluster solutions for all dissimilarity matrices, except for the matrix attached to the P class of arguments. For the P class, the fuzzy cluster algorithm was unable to converge with $ME \leq 1.2$, and with higher exponents, membership coefficients were all close to equiprobable membership (i.e. close to $1/k$). We conclude that there is no appreciable clustering of semantic roles in the P class and exclude it from further analysis. This leaves us with the four classes S, A, G, and T. For these, Figure 3 plots the Dunn statistic as measures of crispness of the fuzzy clustering.

Based on the plots in Figure 3, we performed fuzzy cluster analyses with 3 clusters for S, 6 clusters for A and G each,¹² and 4 clusters for T because these represent the first local maxima or the point where the curve begins to flatten out. The results are shown in Figures 4 and 5. We first discuss the results for the ‘subject’ classes (S and A) and then focus on the results for the ‘object’ roles (G and T) of three-argument predicates.

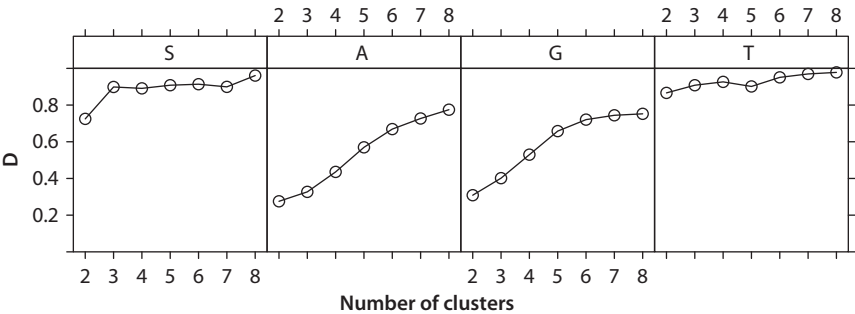


Figure 3. Dunn statistic (D) per number of clusters for each argument class

5.1 S and A classes of semantic roles

The S class (top graph in Figure 4) shows a relatively clear separation between EXPERIENCER roles on the right-hand side of the graph (Cluster 1, colored red) and UNDERGOERS OF BODY PROCESSES on the left-hand side (Clusters 2 and 3). The best exemplars of experiencer roles are located at the bottom of the graph (e.g. *S-feel_like_laughing*, *S-feel_ashamed*, *S-feel_fear*, with membership coefficients higher than .99), but they are closely followed by other roles that are distributed all over (e.g. *S-feel_thirsty*, *S-be_surprised*, *S-feel_cold*, etc., with membership coefficients higher than .98). From a conceptual point of view, no specific structures emerge within the experiencer cluster. For example, no division emerges between emotions and sensations, or between experiences that affect the whole body and experiences that are more specialized. The NeighborNet suggests a continuum within the cluster (so that path lengths between adjacent roles are shorter than between more distant roles), but as far as we can tell, there is no conceptual rationale motivating this.

For undergoers of body processes, by contrast, the fuzzy cluster analysis shows a relatively clear subdivision: Cluster 2 (green) appears to combine processes that are more open to voluntary intervention and control than the processes in Cluster 3 (dark blue). The prototype roles (with membership coefficients higher than .99) for the better controllable processes are *S-urinate*, *S-defecate*, *S-laugh*, *S-shed_tears*, and *S-sleep*. The cluster of less well controllable processes is centered around the roles *S-shiver* and *S-fall*. The roles *S-belch*, *S-be_crazy*, *S-wake_up*, and *S-grow* also belong to the prototype range of less well controllable processes (membership coefficients above .99), but with data support from only two languages each (as shown by the small font size). While the subdivision between less vs. better controllable body processes is well supported by the fuzzy cluster analysis (cf. the high Dunn statistic for a three-cluster analysis of the S class in Figure 3), the NeighborNet also makes clear that there is a continuum between the two, making up an overall cluster of undergoers of body processes. Note that this cluster also contains a number of predicates which one would traditionally analyze as involving themes (*swim*, *grow*, *appear*, *fall*, etc.), but there is no evidence that these constitute a special subcluster.

The experiencer role in the S class is partially replicated in the A class, but the support for this is weaker (in line with generally lower crispness of the clustering, as shown in Figure 3): experiencers tend to be grouped as Cluster 1 (red, on the right-hand side of the bottom graph in Figure 4), centered around *A-feel_irritated_w*, *A-enjoy*, and *A-annoy* (with membership coefficients above .99). However, other roles related to emotions and sensations are scattered across other clusters (e.g. *A-be_tasty_to*, *A-fear*, *A-envy*, etc.). Given the low numbers of

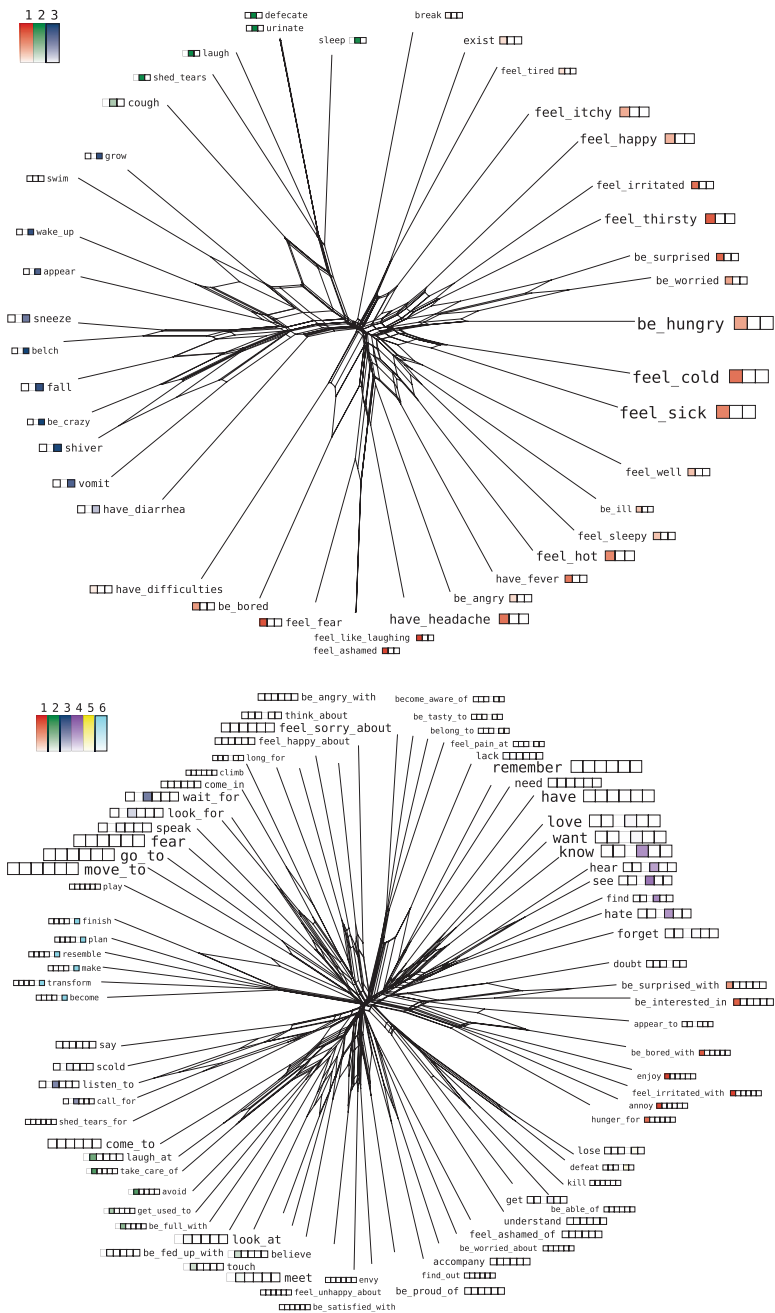


Figure 4. NeighborNet and fuzzy clustering of predicate-specific roles in non-default case assignment to S (top) and A (bottom) arguments. (Roles are represented by predicates, but these are meant to refer to the respective arguments, e.g. break in the S class refers to ‘that which breaks.’)

mentionings (cf. the small font size), however, neither the experiencer cluster nor the exceptions to it should be accorded too much weight in the overall interpretation of the results. The evidence here is clearly weak.

The A class shows a few other tentative clusters that are supported both by the NeighborNet and the fuzzy cluster analysis: one of these is Cluster 4 (purple, on the right-hand side of the graph), which combines A roles in COGNITION AND PERCEPTION, centered around the prototype A-see (with a membership coefficient of .98) and immediately followed by A-find, A-know, and A-hate (membership coefficient of .97). The cluster does not extend to what appears to be related roles such as A-think_about or A-believe, which more often group with other roles in our data. Another cluster with relatively good support is Cluster 6 (light blue, on the left-hand side of the graph), which mostly combines roles involved in TRANSFORMATIONS (A-transform, A-finish, A-make, A-plan, A-become and — not really fitting here, A-resemble, all with a membership coefficient higher than .99). Finally, Cluster 3 (dark blue, upper left) tentatively suggests a cluster of roles involved in COMMUNICATION, with the best representatives being A-listen_to (membership coefficient .97) and A-call_for (.96); A-wait_for shows a similar membership coefficient (.97), but the NeighborNet suggests that the role is also very similar to other roles outside Cluster 3. Also note that the ‘transformer’ and ‘communicator’ roles are supported only by relatively few languages and can only be taken as highly tentative.

5.2 G and T classes of arguments

Like in the A class, the G class of arguments does not show very crisp clustering (cf. the low Dunn statistic in Figure 3), and this is confirmed by the NeighborNet which suggests similarly long distances between most roles (cf. the top graph in Figure 5). But some trends in the analysis are detectable nevertheless. Cluster 1 (red, top right) appears to be clustered around SOURCE locations, specifically around G-get_from and G-buy_from (with membership coefficients of .99 and .98, respectively). Another potential trend is Cluster 2 (green, lower right), with GOALS OF SPATIAL TRANSFER (G-give and G-send_to) at the center. However, the evidence for such a cluster is weak: even the central roles reach membership coefficients of only .90 and .89 and the NeighborNet suggests continuity with G class arguments of COMMUNICATION (*say, ask, teach, explain to*) and TRANSACTION verbs (*sell, borrow from, lend to*) in what is placed in Cluster 5 (yellow, just above Cluster 1) by the fuzzy cluster analysis.

The only other cluster that is supported by closeness in the NeighborNet analysis is Cluster 4 (purple, top left in the graph). This cluster is centered around G-quarrel_with and G-pay_with_for (each with .99 membership coefficients), but it is unclear whether there are any systematic conceptual grounds that would motivate this cluster as distinct from the roles in other parts of the graph.

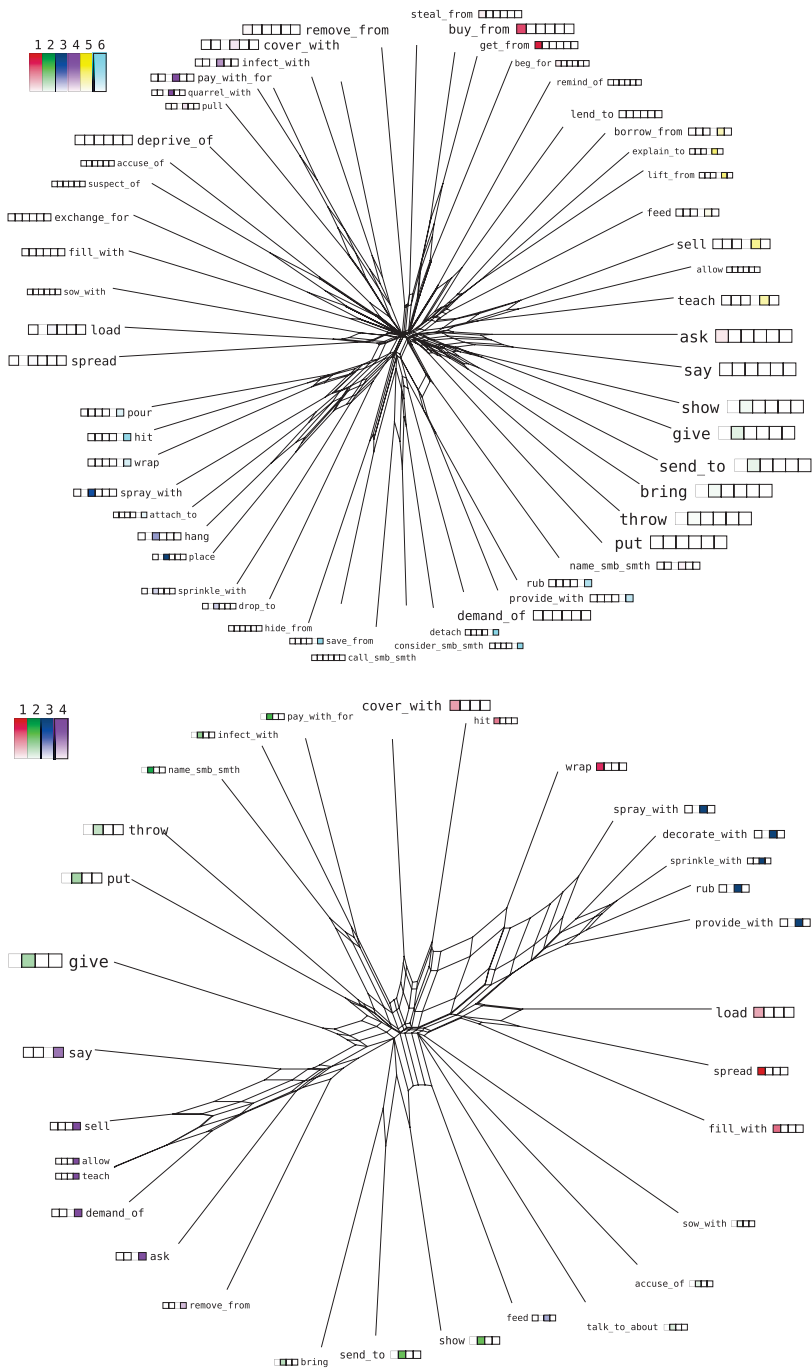


Figure 5. NeighborNet and fuzzy clustering of predicate-specific roles in non-default case assignment to G (top) and T (bottom) arguments. (Labeling conventions as in Figure 4)

The trend towards a transfer-based cluster around G-give and G-send_to is not paralleled by the data from T class arguments (cf. the bottom graph in Figure 5). T-give and T-send_to appear in Cluster 2 (green) but not as prototypes, and the NeighborNet analysis shows that they are at almost maximum distance from each other (T-give appears in the upper left area of the graph while T-send_to is placed at the bottom). What seems more relevant for the overall structure in the T class is a continuum between THEME-OF-MOTION meanings starting in the middle of the graph (Cluster 2, green, both upper and lower side) and ending on the right-hand side with roles covering material and things that become ATTACHED to something (Cluster 1 and 3, red and blue). The only apparent exception is T-hit (top side), but the conceptual distance is not too far in fact: an instrument that one uses for hitting becomes attached to an object if only perhaps briefly while, say, a cover or load tends to stay in place once it is applied. However, we note that the area of attached items does not cluster too well, spread as it is between two of the clusters suggested by the fuzzy cluster analysis. Also, themes of transfer motion appear on both the upper and lower side in the middle of the NeighborNet, showing more diversity than the fuzzy cluster analysis suggests by grouping roles in Cluster 2 (green).

The only cluster that is consistently separate across both the Fuzzy Cluster and the NeighborNet analysis in the T class is Cluster 4 (purple, lower left side). This cluster is mostly about SPEECH arguments ('what is communicated') and is centered around T-allow, T-teach, T-demand_of (all with membership coefficients above .99). Some roles do not fit, however, such as T-sell (which has a membership coefficient of .99 as well) and some verbs of saying are placed elsewhere (e.g. T-accuse_of, T-name).

6. Discussion

This study is based on surveying lists of predicates associated with non-default case assignment, as found in descriptive grammars and dictionaries. As noted in Section 3, this method of data collection necessarily brings with it a certain amount of uncertainty. However, there is evidence that, despite this, the data are sufficiently clear: we do find some clear clustering effects in some areas but not in others. If the data were too noisy or too skewed (e.g. because some predicate meanings were systematically underreported in grammars), one would have expected much fewer or much more (and crisper) clustering than we actually found. This suggests that the results allow further interpretation and comparison with traditional expectations about semantic role clusters.

There are a few areas where our analysis confirms traditional expectations. The best fit with traditional role notions is found in the class of S arguments. Here,

a broad distinction emerges between EXPERIENCERS as opposed to UNDERGOERS OF BODY PROCESSES, the latter subdivided into more vs. less controllable processes (e.g. *urinate* vs. *shiver*). The experiencer cluster is weakly supported by the A class of arguments as well. While the S class does not suggest any further subdivisions of experiencers into, say, emotions vs. sensations, the A class shows a separation of experiencers vs. COGNIZERS/PERCEIVERS, although the signal here is not very strong. Another traditional role that can be detected is the spatial SOURCE role complex among G arguments.

The other argument classes do not show clusterings reminiscent of traditional semantic role types. Most importantly, the analysis of the T class of arguments fails to support a clear distinction between instruments and themes. What emerges instead is a loosely structured continuum ranging from THEMES of transfer to ATTACHED ITEMS. While unexpected from the point of view of a *a priori* theorizing, this result is in line with the observation that it is often difficult to tell apart instruments and themes. For example, the T argument of verbs like ‘to fill’ can be conceived both as a theme that is moved into a container or as an instrument with which one changes the state of the container. What seems more relevant is a notion of attachment, whereby items end up in an attached or otherwise fixed state.

More generally, it seems that spatial transfer is less relevant for three-argument predicates than traditionally assumed. There is a cluster of GOALS OF SPATIAL TRANSFER among G arguments (with the goals of *give* and *send to* at the center), but as noted above, the cluster is not well separated from verbs of communication (*say*) and transaction (*sell*). This can mean that, despite widespread assumptions, spatial transfer is not a prototypical concept for three-argument predicates. However, our finding can also mean that spatial transfer is only relevant as a prototype outside non-default case assignment. This second possibility would require that the opposite of non-default case assignment, i.e. ‘basic’ case frames, is not conceived of as a default in our sense (cf. Section 2), but that ‘basic’ case frames are conceived of as positively defined semantic categories, organized around prototypes. Evidence for such an interpretation cannot come from our study (since it is limited to non-basic case frames), but requires in-depth work on the semantic fine-structure of basic case frames in three-place predicates.

Our study also suggests a number of role clusters that are not canonically assumed in theories of semantic roles. The most important concept here is that of SPEECH/CONCEPT. A role cluster of communication items (e.g. T-allow or T-demand_of) emerges as a relatively well supported cluster in the T-class, and there is tentative and weak evidence for a cluster of ‘communicator’ roles in the A class (e.g. A-listen_to or call_for).

In general, clustering is weaker among the A and G classes of roles than among the S and T classes. It seems that when languages assign non-default cases to A

and G arguments, they do so in ways that cannot be easily predicted by predicate semantics. For the A class, this may perhaps be explained by assuming that case markers on A arguments mostly serve the discriminatory rather than indexing function of case (Comrie 1978, 1981, Song 2001), i.e. these case markers simply keep A noun phrases distinct from P noun phrases and do not index any semantic subgroups of roles. This explanation fits with the observation that the P argument class does not show any role clustering at all. However, the absence of statistical signals can also be caused by the limited size of our dataset and we do not want to push any explanation here too far. Also, we note that the distinction between discriminatory and indexing functions is not as straightforward as traditionally assumed (Arkadiev 2008).

For the G class, one possible explanation comes from the fact that these roles are commonly associated with various locative cases and that locative cases (e.g. various subtypes of allatives) tend to easily extend and reduce their semantic ranges over time for reasons that have little to do with predicate semantics and more with the conceptualization of space. In order to substantiate this, one would now have to compare the results from G class clusters with clusters based on the spatial meanings of the relevant case markers. We leave this for future research.

7. Conclusions

This study has looked into patterns of non-default case assignment as possible evidence for semantic role clustering. In order to do so, we have combined methods that allow for the detection of categories that are organized around prototypes as well as the detection of continua and cross-cluster similarities. While the dataset is limited, it does suggest that some of the traditionally assumed role clusters play a cross-linguistically significant role: EXPERIENCERS, UNDERGOERS OF BODY PROCESSES, and COGNIZERS/PERCEIVERS in one- and two-place predicates; SOURCES and transmitted SPEECH/CONCEPT in three-place predicates. Our study provides no support for a categorial distinction between themes and instruments, but weak trends for a continuum between THEMES of transfer and ATTACHED ITEMS and, as a possibly distinct cluster, GOALS of transfer. This suggests that in these areas, languages treat predicate-specific roles in a much more varied way than traditional theories would lead us to expect.

In the introductory sections, we contrasted our approach with alternatives that are based on pre-compiled predicate lists for which case frames can be surveyed and compared. Both approaches come with problems — our approach with the problem of noise in the data (individual predicates may be forgotten when non-default case frames are described in a language or compositional patterns may be

misanalyzed as lexicalized complex predicates), list-based approaches with the problem of *a priori* skewing of results (specific predicate types may be overrepresented or underrepresented when drawing up a list and universal semantic analyses may be forced upon languages, artificially reducing diversity). Ultimately one will want to compare the results of both approaches. Hypotheses of universally relevant semantic role clusters need to pass muster across methods. Finally, any such hypothesis should be taken to be tentative before it is subjected to a diachronic and area-by-area analysis that probes into the dynamics of role clusters over time and space.

Notes

* Versions of this paper were presented at the workshop on role complexes in Zurich, 4–5 April 2011, and the conference on non-canonically case-marked subjects in Reykjavik, 4–8 June 2012. We thank both audiences for helpful comments and questions. Author contributions: BB, AWM, and TZ conceived and designed the study. BB and TZ conducted the statistical analysis. LB and AWM did most of the data analysis and database coding. TZ did most of the data aggregation work. BB and AWM wrote the paper. We thank Kevin Bäscher for help in data collection, Lukas Wiget for help in preparing the graphs, and Sergey Say for comments on an earlier draft. We also thank three anonymous reviewers and an anonymous editorial board member for helpful suggestions and requests for clarification.

1. We use the term ‘case assignment’ in the broad sense of a paradigmatic contrast in the shape of noun phrases that distinguishes their roles as arguments of a predicate, including affixes, tone oppositions, adpositions, particles, morphological zeros in opposition to overt devices, etc.
2. Glossing follows the Leipzig Glossing Rules; ‘V’, ‘J’, and ‘B’ denote genders.
3. In addition, we note that it can be very difficult to decide whether a specific case assignment is motivated by some sub-structure of the lexical predicate. The answer will often depend on the precise etymology of the expression and on the question to what extent speakers still have access to this sub-structure.
4. This is the approach taken by the Leipzig Valency Class Project (Comrie & Malchukov In press, <http://www.eva.mpg.de/lingua/valency>) and the Valency Project at the Russian Academy of Sciences, St. Petersburg (Say 2011).
5. Although in principle, class size and productivity might be in conflict, in practice they seem to coincide in determining one most common predicate class. For instance, Barðdal found a correlation between syntactic productivity and the size of predicate classes in the lexicon, as well as between the size of predicate classes and text frequency in Icelandic (Barðdal 2008: 60, 172f.). As the information on the number of predicates belonging to a particular lexical predicate class is relatively readily available for the languages surveyed, we chiefly rely on this criterion in our survey.
6. In a more extensive survey, Witzlack-Makarevich (2011) indeed finds no single language where case assignment would ever split one-argument predicates into classes that differ by a

margin of less than 10% in their sizes. For case studies, see for example, Merlan (1985), Holisky (1987), or Pustet (2002). For a comparative lexicon investigation, see Nichols (2008).

7. The definition of the G vs. T contrast deviates from the one in Bickel et al. (2010) and Witzlack-Makarevich (2011), partly in response to an insightful critique in Schikowski (2013). For a recent comparison of various approaches to thematic roles including some critique of Dowty's approach, see Croft (2012). Note that we differ from Dowty (1991) in that we set up A and P not as independent but as relative prototypes, following Primus (1999, 2006).

8. The database and the statistical analyses reported below are available as an online appendix at DOI 10.1075/sl.38.3.03bic.additional.

9. For a diachronic study, one would have to have much denser sampling in large families.

10. Our analyses can be inspected and checked in the online appendix.

11. Detailed inspection of frequencies is possible in the online appendix.

12. For A and G we also considered analyses with more clusters, but the NeighborNet analyses printed below show that any further clusters are so close to each other that the partition would seem an artifact of the method, with limited empirical interest.

References

- Arkadiev, Peter M. 2008. Differential argument marking in two-term case systems and its implications for the general theory of case marking. In Peter de Swart & Helen de Hoop (eds.), *Differential subject marking*, 151–171. Dordrecht: Springer.
- Barddal, Johanna. 2008. *Productivity: Evidence from case and argument structure in Icelandic*. Amsterdam: Benjamins.
- Bickel, Balthasar. 2008/2011. *Grammatical relations typology*. In Jae Jung Song (ed.), *The Oxford handbook of language typology*, 399–444. Oxford: Oxford University Press.
- Bickel, Balthasar & Johanna Nichols. 2009. Case marking and alignment. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 304–321. Oxford: Oxford University Press.
- Bickel, Balthasar, Manoj Rai, Netra Paudyal, Goma Banjade, Toya Nath Bhatta, Martin Gaenszle, Elena Lieven, Iccha Purna Rai, Novel K. Rai & Sabine Stoll. 2010. The syntax of three-argument verbs in Chintang and Belhare (Southeastern Kiranti). In Andrej Malchukov, Martin Haspelmath & Bernard Comrie (eds.), *Studies in ditransitive constructions: A comparative handbook*, 382–408. Berlin: Mouton de Gruyter.
- Bresnan, Joan & Jonni Kanerva. 1989. Locative inversion in Chichewa: A case study in factorization in grammar. *Linguistic Inquiry* 20. 1–50.
- Bryant, David & Vincent Moulton. 2004. Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21. 255–265. DOI: 10.1093/molbev/msh018
- Butt, Miriam. 2008. Case in Lexical-Functional Grammar. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford handbook of case*, 59–71. Oxford: Oxford University Press.
- Comrie, Bernard. 1978. Ergativity. In Winfred P. Lehmann (ed.), *Syntactic typology: Studies in the phenomenology of language*. Austin: University of Texas Press.

- Comrie, Bernard. 1981. *Language universals and linguistic typology*. Chicago: University of Chicago Press.
- Comrie, Bernard. 2005. Alignment of case marking. In Martin Haspelmath, Matthew S. Dryer, David Gil & Bernard Comrie (eds.), *The world atlas of language structures*, 398–405. Oxford: Oxford University Press.
- Comrie, Bernard & Andrej Malchukov (eds.). In press. *Valency classes: A comparative handbook*. Berlin: de Gruyter Mouton.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198299554.001.0001
- Croft, William. 2012. *Verbs: Aspect and causal structure*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199248582.001.0001
- Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511993473
- Dixon, R. M. W. 1994. *Ergativity*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511611896
- Dowty, David R. 1991. Thematic proto-roles and argument selection. *Language* 67. 547–619. DOI: 10.1353/lan.1991.0021
- Dryer, Matthew S. 1989. Large linguistic areas and language sampling. *Studies in Language* 13. 257–292. DOI: 10.1075/sl.13.2.03dry
- Dryer, Matthew S. 1997. Are grammatical relations universal? In Joan Bybee, John Haiman & Sandra A. Thompson (eds.), *Essays on language function and language type dedicated to T. Givón*, 117–143. Amsterdam: Benjamins.
- Haspelmath, Martin. 2001. Non-canonical marking of core arguments in European languages. In Aleksandra Y. Aikhenvald, R. M. W. Dixon & Masayuki Onishi (eds.), *Non-canonical marking of subjects and objects*, 53–84. Amsterdam: Benjamins. DOI: 10.1075/tsl.46.04has
- Haspelmath, Martin. 2011. On S, A, P, T, and R as comparative concepts for alignment typology. *Linguistic Typology* 15. 535–567. DOI: 10.1515/LITY.2011.035
- Holisky, Dee A. 1987. The case of the intransitive subject in Tsova-Tush (Batsbi). *Lingua* 71. 103–132. DOI: 10.1016/0024-3841(87)90069-6
- Huson, Daniel H. & David Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23. 254–267. DOI: 10.1093/molbev/msj030
- Kaufman, Leonard & Peter J. Rousseeuw. 1990. *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.
- Maechler, Martin, Peter J. Rousseeuw, Anja Struyf & Mia Hubert. 2005. cluster: Cluster analysis basics and extensions. R package, <http://www.R-project.org/>.
- Malchukov, A. 2005. Case pattern splits, verb types and construction competition. In Mengistu Ambember & Helen de Hoop (eds.), *Competition and variation in natural languages: the case for case*, 73–118. Amsterdam: Elsevier. DOI: 10.1016/B978-008044651-6/50006-9
- Malchukov, Andrej, Martin Haspelmath & Bernard Comrie. 2010. Ditransitive constructions: A typological overview. In *Studies in ditransitive constructions: a comparative handbook*, 1–35. Berlin: de Gruyter Mouton.
- Merlan, Francesca. 1985. Split intransitivity: Functional oppositions in inflections. In Johanna Nichols & Anthony C. Woodbury (eds.), *Grammar inside and outside the clause: Some approaches to theory from the field*, 324–62. Cambridge: Cambridge University Press.
- Molochieva, Zarina. 2010. *Tense, aspect, and mood in Chechen*. Leipzig: University of Leipzig PhD dissertation.

- Nichols, Johanna. 2008. Why are stative-active languages rare in Eurasia? A typological perspective on split subject marking. In Mark Donohue & Søren Wichmann (eds.), *The typology of semantic alignment*, 121–139. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199238385.003.0005
- Nichols, Johanna, Alena Witzlack-Makarevich & Balthasar Bickel. 2013. The autotyp genealogy and geography database: 2013 release. Electronic database, <http://www.uzh.ch/spw/autotyp>.
- Onishi, Masayuki. 2001. Introduction: Non-canonically marked subjects and objects: Parameters and properties. In Aleksandra Y. Aikhenvald, R. M. W. Dixon & Masayuki Onishi (eds.), *Non-canonical marking of subjects and objects*, 1–52. Amsterdam: Benjamins. DOI: 10.1075/tsl.46.03oni
- Pandharipande, Rajeshwari V. 1997. *Marathi*. London: Routledge.
- Primus, Beatrice. 1999. *Cases and thematic roles*. Tübingen: Niemeyer. DOI: 10.1515/9783110912463
- Primus, Beatrice. 2006. Mismatches in semantic role hierarchies and the dimensions of role semantics. In Ina Bornkessel, Matthias Schlesewsky, Bernard Comrie & Angela D. Friederici (eds.), *Semantic role universals and argument linking: Theoretical, typological and psycholinguistic perspectives*, 53–87. Berlin: Mouton de Gruyter.
- Pustet, Regina. 2002. Split intransitivity revisited: Comparing Lakota and Osage. *International Journal of American Linguistics* 68(4). 381–427. DOI: 10.1086/466499
- R Development Core Team. 2013. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, <http://www.r-project.org>.
- Rokach, Lior. 2010. A survey of clustering algorithms. In Oded Maimon & Lior Rokach (eds.), *Data mining and knowledge discovery handbook*, 269–298. New York: Springer [second edition].
- Sava, Graziano. 2005. *A grammar of Ts'amakko*. Cologne: Koppe.
- Say, Sergey. 2011. Nekanoničeskoe markirovanie aktantov mnogomestnyx predikatov: opyt kvantitativno-tipologičeskogo issledovanija. *Acta Linguistica Petropolitana* 7. 424–430 [http://iling.spb.ru/pdf/alp/alp_VII_3.pdf].
- Schikowski, Robert. 2013. *Object-conditioned differential marking in Chintang and Nepali*. Zurich: University of Zurich PhD dissertation.
- Song, Jae Jung. 2001. *Linguistic typology: Morphology and syntax*. Harlow: Longman.
- Tsunoda, Tasaku. 1985. Remarks on Transitivity. *Journal of Linguistics* 21. 385–396. DOI: 10.1017/S002226700010318
- Tsunoda, Tasaku. 2004. Issues in case-marking. In Peri Bhaskararao & Karumuri Venkata Subbarao (eds.), *Non-nominative subjects*, vol. 2, 197–208. Amsterdam: Benjamins. DOI: 10.1075/tsl.61.11tsu
- Van Valin, Robert D., Jr. 2005. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511610578
- Van Valin, Robert D., Jr. & David P. Wilkins. 1996. The case for 'effector': Case roles, agents, and agentivity revisited. In Masayoshi Shibatani & Sandra A. Thompson (eds.), *Grammatical constructions*, 289–322. Oxford: Oxford University Press.
- Witzlack-Makarevich, Alena. 2011. *Typological variation in grammatical relations*. Leipzig: University of Leipzig PhD dissertation.